



I lavori raccolti in questa collana hanno avuto origine nell'ambito del progetto di ricerca dell'ISTAT «Le statistiche dei consumi privati nel sistema statistico nazionale» e del progetto di ricerca MURST 40% «La misura dei consumi privati: uno studio sull'accuratezza, coerenza e qualità dei dati». Al progetto di ricerca hanno partecipato i ricercatori dell'ISTAT e dei seguenti Dipartimenti e Istituti universitari:

- Dipartimento di Scienze Statistiche, Bologna
- Dipartimento di Contabilità Nazionale, Roma
- Dipartimento Statistico, Firenze
- Istituto di Statistica e Matematica, Istituto Universitario Navale, Napoli
- Dipartimento di Scienze Statistiche, Perugia
- Istituto di Statistica, Messina.

**PROPOSTE ALTERNATIVE ALLA ROTAZIONE
COMPLETA NELL' INDAGINE SUI CONSUMI
DELLE FAMIGLIE**

Michele Lalla*

Rapporto di ricerca n.5

CON PRI - La misura dei consumi privati



* Istituto di Statistica sociale e scienze demografiche e Biometriche,
Università di Palermo

Dipartimento di Scienze Statistiche "Paolo Fortunati"
dell' Università degli Studi di Bologna
Novembre 1992

BIBL. DIP. DI SCIENZE STATISTICHE

Statistica
Q.2

UNIVERSITÀ DEGLI STUDI DI BOLOGNA

8772

INDICE

1. Introduzione	p.	5
2. Valutazione dei diversi metodi di indagine	"	6
2.1. Metodi di indagine dei fenomeni nel tempo	"	7
2.2. Caratteristiche di alcuni metodi di indagine	"	8
2.3. La rilevazione dei consumi: considerazioni su alcuni metodi di indagine	"	11
3. Aspetti caratteristici di un piano di campionamento per la rilevazione dei consumi: l'indagine panel e rotata	"	14
3.1. Periodo di registrazione delle spese	"	15
3.2. Stratificazione delle unità di primo stadio	"	15
3.3. L'indagine panel limitata nel tempo	"	16
3.4. L'indagine rotata	"	19
3.5. La dinamica della popolazione obiettivo	"	22
4. Metodi di analisi dei dati raccolti con indagini panel	"	24
4.1. Metodo campionario: stime di livello	"	25
4.2. Metodi di analisi multivariata	"	27
4.3. Metodi di analisi delle durate	"	29
5. Conclusioni	"	32
<i>Note</i>	"	34
<i>Riferimenti bibliografici</i>	"	36

1. Introduzione^(*)

L'ampiezza e la complessità del fenomeno del consumo e dell'analisi dei bilanci delle famiglie, nonché la molteplicità delle fasi dell'indagine, richiedono un'approfondita analisi teorica e un coordinamento efficiente del processo di raccolta e di elaborazione dei dati. La precisazione delle finalità e dei metodi di rilevazione della spesa è essenziale per individuare le variabili rilevanti dello studio, per scegliere il metodo di indagine, e per definire il piano di campionamento. La scelta tra i diversi metodi di indagine, poi, è determinata anche dalla complessità della struttura operativa necessaria per condurre l'indagine, dalle competenze richieste ai singoli componenti del gruppo di lavoro, dalla qualità e dal tipo di dati da rilevare, dalle possibilità di stima e di modelli che si possono realizzare, dall'analisi statistica da condurre, perché questi elementi concorrono a definire i costi.

La revisione dell'indagine sui consumi e sui bilanci delle famiglie deve interessare, quindi, tutte le fasi del processo di rilevazione: finalità e obiettivi, piano di campionamento, organizzazione operativa, strumenti di rilevazione, eccetera. L'analisi dei molteplici aspetti dell'indagine attuale condotta dall'Istat, delle difficoltà che presenta nella sua realizzazione, e delle soluzioni possibili per migliorare la strategia di rilevazione sono state sviluppate nell'ambito del progetto di ricerca *CON.PRI.*: "Le statistiche dei consumi privati nel Sistema Statistico Nazionale" (Filippucci e Marliani 1992a, b; Mantegazza e Tassinari 1992).

L'argomento trattato nel seguito concerne la valutazione dei diversi metodi di indagine utilizzabili per la rilevazione dei consumi privati e dei bilanci delle famiglie con riferimento alla possibilità di applicare metodi diversi dall'indagine ripetuta, che attualmente è adottata dall'Istat, per stimare i totali, gli indici di posizione, gli indici di variabilità di alcune caratteristiche essenziali del consumo; per analizzare la dipendenza tra i caratteri e i cambiamenti rispetto alle variabili individuali e famigliari, al reddito, al risparmio, alla regione, ai segmenti economici e sociali, e così via. In

^(*) Si ringrazia Andrea Giommi per aver letto il lavoro e per le preziose osservazioni che mi ha comunicato. Si ringrazia particolarmente Carlo Filippucci per la lettura attenta e puntigliosa e per gli utili suggerimenti che hanno contribuito a migliorare il testo in molti punti. Rimane valida la solita avvertenza: responsabile di qualunque errore e omissione è solo l'autore.

particolare, si esamineranno alcuni schemi di campionamento delle indagini panel e rotate perché le loro potenzialità sono più vaste di quelle dell'indagine ripetuta e la loro applicazione investe, comunque, tutti gli aspetti dell'indagine: gli strumenti di rilevazione, la struttura organizzativa, lo svolgimento delle operazioni sul campo (formazione dei rilevatori, raccolta dei dati, controllo della loro qualità, eccetera), la costruzione di modelli relativi al comportamento dei consumatori (Sudman e Ferber 1979; Deaton 1985; Blundell 1986, 1988). Inoltre, si presenterà anche una breve panoramica dei procedimenti di stima dei dati raccolti con le indagini panel e rotate.

L'esposizione procede come segue. Nel paragrafo 2 si descrivono le diverse tecniche che si adoperano per conseguire gli obiettivi specifici delle indagini campionarie relative a fenomeni che presentano una dimensione temporale. In particolare, si esamina la possibilità di utilizzare l'indagine panel nella rilevazione dei consumi e dei bilanci di famiglia. Nel paragrafo 3 si illustrano gli elementi caratteristici del piano di campionamento dell'indagine sui consumi delle famiglie e alcuni schemi di indagine longitudinale per il panel limitato nel tempo e rotato. Nel paragrafo 4 si considerano i metodi di stima, gli stimatori, e i modelli usati più frequentemente nell'analisi dei dati raccolti con le indagini panel o rotate. Il paragrafo 5 riporta le principali conclusioni.

2. Valutazione dei diversi metodi di indagine

Il consumo è un fenomeno che si manifesta ed evolve nel tempo. Per analizzarlo è necessario, quindi, introdurre la dimensione temporale nello schema di campionamento perché le caratteristiche e la composizione della popolazione di riferimento possono cambiare, compromettendo il requisito della comparabilità dei dati rilevati nelle varie fasi (Kish 1986). Un'ampia panoramica sugli aspetti teorici del consumo, sui problemi di misurazione, e sulle strategie di rilevazione, si trova in Filippucci e Marliani (1992a,b).

La stima dei parametri della popolazione al variare del tempo, tramite campione, fu analizzata in principio da Jessen (1942), Yeats (1949), e Patterson (1950). Traendo spunto dal sistema di rilevazione della popolazione corrente adottato dal Bureau of Census degli Stati Uniti d'America, successivamente furono condotti diversi studi sugli schemi di rotazione da Hansen e altri (1955), Eckler (1955), Rao e Graham (1964), Gurney e Daly (1965).

Gli obiettivi che si conseguono con le indagini dei fenomeni nel tempo, in genere, sono (Duncan e Kalton 1987; Kasprzyk e altri 1989):

- (a) stima dei parametri della popolazione relativi a tempi distinti;
- (b) stima dei valori medi nel tempo dei parametri della popolazione¹;
- (c) stima della variazione netta (*net changes*);
- (d) stima dei cambiamenti individuali (*gross changes*);
- (e) cumolazione dei dati individuali nel tempo;
- (f) cumolazione dei campioni nel tempo;
- (g) osservazione di eventi che accadono in un periodo fissato².

2.1. Metodi di indagine dei fenomeni nel tempo

Il metodo dell'*indagine ripetuta o periodica* consiste nell'effettuare osservazioni o misure simili su campioni distinti³ estratti, in momenti diversi del tempo, da una popolazione equivalente che è univocamente definita⁴. L'indagine ripetuta è la più semplice da realizzare e permette di conseguire gli obiettivi: (a), (b), (c), (f)⁵.

Il metodo dell'*indagine panel*, proposto da Lazarsfeld e Fiske (1938), richiede la rilevazione delle stesse misure sullo stesso campione a momenti diversi del tempo, ovvero è un'indagine ripetuta nel tempo sulle stesse unità⁶. Con l'indagine panel si conseguono gli obiettivi (c), (d), (e), (g), ma la sua realizzazione è assai problematica perché il campione è soggetto al logoramento e alla conseguente perdita di rappresentatività, i dati raccolti possono essere distorti dal condizionamento dei soggetti nel fornire le risposte o dall'onere della continuità nel tempo dell'intervista⁷.

Il metodo dell'*indagine rotata o panel rotante* osserva gli stessi elementi solo per un periodo limitato: a ogni fase una quota di elementi, $Q = 1 - P$, lascia il campione, e una quota di nuovi elementi, pari a Q , viene aggiunta⁸. Si possono conseguire gli obiettivi: (a), (b), (c); mentre (d), (e), (f), (g) sono penalizzati dalla breve permanenza degli elementi nel campione. I piani di campionamento si distinguono per l'ammontare di P e per il tipo di rotazione tra i periodi, ma se $P = 1$ si ha una indagine panel e se $P = 0$ si ha una indagine ripetuta.

Il metodo dell'*indagine split panel* utilizza un campione suddiviso in due componenti distinte e "indipendenti": una panel e l'altra ripetuta o rotata (Kish 1983, 1986); infatti, il termine *split* significa "diviso o spaccato o

scisso". La componente indagine panel consente di ottenere gli obiettivi (c), (d), (e), (g). La componente indagine ripetuta permette di conseguire gli obiettivi (a), (b), (c), (f); di accertare gli ingressi e le uscite che avvengono nella popolazione; di controllare la distorsione nell'indagine panel data dal condizionamento, dalla perdita dei soggetti, e dal logoramento del campione.

Il metodo dell'*indagine trasversale con domande retrospettive* o semplicemente *retrospettiva* si caratterizza per la raccolta, su ogni unità, di dati riferiti sia al momento in cui si realizza la misura, sia a momenti precedenti; si basa sul ricordo dei soggetti intervistati per ricostruire il flusso di informazioni nel tempo. I dati possono presentare notevoli errori di tipo non campionario (Sudman e Brandburn 1973; Silberstein 1973; Moss e Goldstein 1979; Cannell, Miller, e Oksenberg 1981; Mathiowetz e Duncan 1984) perché gli intervistati possono dimenticare un evento ("errore di memoria") oppure non ricordare con precisione quando sia avvenuto e la sua entità ("errore telescopico"). A questi problemi di deve aggiungere la distorsione dovuta alla perdita delle unità che sono uscite dalla popolazione prima della data dell'indagine⁹. Malgrado queste limitazioni, si possono conseguire gli obiettivi (d), (e), (g), e, con difficoltà, (a), (b), (c), (f).

2.2. Caratteristiche di alcuni metodi di indagine

L'indagine retrospettiva è stata richiamata per completezza, ma è evidente che non si può utilizzare nel caso della rilevazione dei consumi delle famiglie per le sue caratteristiche e per la natura delle informazioni da rilevare. Solo alcuni dati, come le spese per il telefono, per l'energia elettrica, per l'acqua e il gas, possono essere raccolti anche con domande retrospettive, purché il periodo di riferimento sia breve.

L'indagine panel offre potenzialità enormi per lo studio del fenomeno del consumo. Tra l'altro, consente di sviluppare il concetto di monitoraggio della spesa e dei consumi perché i dati raccolti, secondo questa tecnica, permettono di esaminare l'andamento temporale dei consumi, l'evoluzione dei comportamenti dei consumatori, le componenti di stabilità o di mutamento delle unità statistiche; per esempio: quali tipologie famigliari aumentano le spese per i beni semidurevoli e quali le diminuiscono; quali fattori incidono sull'aumento di spesa delle famiglie; quali condizioni incidono sulle variazioni di spese sanitarie nel ricorso a cure specialistiche; quali eventi causano

la rinuncia alle vacanze; (Duncan e Hill 1985). Inoltre, si possono modellare i comportamenti individuali sia per spiegare e prevedere il loro effetto sul consumo, sia per l'analisi di aspetti teorici che verrebbero trascurati senza la disponibilità di dati microeconomici (Blundell 1988; Deaton 1992). Anche per i beni è possibile approfondire aspetti importanti come il "ciclo di vita" dei prodotti acquistati. L'utilità dell'indagine panel è, quindi, evidente; ma esistono difficoltà nella sua realizzazione che ne limitano l'impiego.

L'indagine ripetuta, invece, permette di osservare solo la variazione netta (per esempio, l'aumento o la diminuzione delle spese per un bene) o l'andamento delle spese nel tempo, le tipologie famigliari che effettuano maggiori spese per i beni semidurevoli, i fattori che incidono sulle spese delle famiglie, il numero di persone che vanno in vacanza, e così via.

Le indagini panel e le indagini ripetute usano criteri antitetici per includere i soggetti nelle fasi successive; pertanto, le difficoltà e le facilitazioni, che ne conseguono dalla loro applicazione, costituiscono una guida nella scelta del metodo di indagine (Sudman 1976; Sudman e Ferber 1979). I problemi e gli svantaggi che derivano dall'uso delle indagini panel sono:

- 1) *auto-selezione iniziale* dei soggetti inclusi nel campione;
- 2) *attrito* o perdita di soggetti per stanchezza, noia, apatia o irreperibilità;
- 3) *non risposta temporanea* dovuta o alla non presenza a casa o al rifiuto di rispondere che può essere più alta dell'attrito (3-6% contro l'1-2%);
- 4) *reattività alla reintervista*, perché l'esperienza di quella passata e l'anticipazione di un'altra da realizzare in futuro possono cambiare il comportamento, le attitudini, e le opinioni dell'intervistato; il fenomeno è detto anche condizionamento, contaminazione, sensibilizzazione, apprendimento, e distorsione del panel;
- 5) *disattenzione nella reintervista* dovuta sia all'intervistato, sia all'intervistatore per accidia, affaticamento, monotonia;
- 6) *mobilità* dei soggetti che rende la loro reperibilità difficile;
- 7) *cambiamenti* subiti nel tempo dalle unità che complicano l'elaborazione dei dati (le separazioni, i divorzi, i matrimoni dei figli, eccetera);
- 8) *mortalità e emigrazione* che riducono l'ampiezza del campione;
- 9) *natalità e immigrazioni* che sono ignorate nelle indagini panel, mentre le indagini ripetute le incorporano automaticamente (Sobol 1959);
- 10) *verifiche e controlli* numerosi occorrono sia per evitare distorsioni e perdita dei soggetti nel campione, sia per ottenere un'alta qualità dei dati, in termini di completezza, consistenza, e coerenza;

- 11) *limiti nell'analisi*, perché il condizionamento dei soggetti e l'onere continuo nel tempo comportano errori di misura e/o omissione di variabili;
- 12) *scelta del periodo di reintervista*, perché la reattività dell'intervistato, (4), potrebbe indicare periodi ottimali che sono in conflitto con le esigenze conoscitive del fenomeno (Cantor 1989).

Si noti che tra le difficoltà citate sono state incluse, per completezza, anche quelle che sono tipiche di ogni indagine campionaria: (7), (8), (9), (10), (11), e (12). Così, per esempio, le verifiche e i controlli, (10), per ottenere dati accurati e non distorti o i limiti nell'analisi, (11), che emergono da errori di misura e/o omissioni di variabili, riguardano un po' tutte le indagini. La loro rilevanza, invece, dipende dal metodo di indagine, dal contesto specifico in cui avviene l'applicazione, e dal tipo di rilevazione che si deve effettuare. I vantaggi delle indagini panel rispetto alle indagini ripetute sono:

- 1) *potenzialità di analisi*, perché i "dati panel" sono più numerosi e più variabili sicché diminuisce la collinearità tra i caratteri esplicativi e aumentano i gradi di libertà e l'efficienza delle stime; i dati individuali consentono di studiare la natura del cambiamento e del comportamento, mentre dalle serie storiche aggregate possono emergere anche comportamenti spuri (Lippi 1985); la varianza nelle osservazioni ha una componente inter-individuale dominante sul resto, che assicura una maggiore robustezza ad alcune stime (Dormont, 1989);
- 2) *rimozione e riduzione dell'errore*, perché si possono introdurre procedure di controllo della coerenza e consistenza dei dati;
- 3) *famigliarità* tra gli intervistati e gli intervistatori, perché spesso facilita il flusso di informazioni dai primi ai secondi in contrasto con l'attrito (2), la reattività (4), e la disattenzione (5);
- 4) *effetto organizzazione*, perché la necessità di disporre di una struttura che funzioni continuamente nel tempo, comporta uno sviluppo di conoscenze e risorse che contribuiscono a migliorare la realizzazione delle diverse fasi dell'indagine;
- 5) *costi minori*, perché l'acquisizione dei dati di base, o che non mutano nel tempo, non si deve ripetere nelle fasi successive alla prima; la conoscenza dei soggetti facilita il contatto in termini di spesa e di tempo; tuttavia, la questione dei costi è controversa e la convenienza non sembra scontata (Kish 1986, 1989; Kalton 1987).

Le difficoltà menzionate, a nostro avviso, non sono preponderanti rispetto ai vantaggi che si hanno nell'uso dell'indagine panel anche perché solo con "dati

panel" si possono analizzare adeguatamente aspetti importanti dei fenomeni che evolvono nel tempo, come il comportamento dei consumatori, la mobilità degli individui, i cambiamenti di stato. Così, gli studi sul consumo e sulla disoccupazione hanno evidenziato l'opportunità di utilizzare microdati per approfondire aspetti che, altrimenti, resterebbero inesplorati (Allison 1982; Tuma e Hannan 1984; Blundell 1988; Dormont 1989; Deaton 1992).

2.3. La rilevazione dei consumi: considerazioni su alcuni metodi di indagine

Le peculiarità dei metodi di indagine esaminati permettono già di svolgere alcune considerazioni in merito al loro uso nella rilevazione dei consumi, che è caratterizzata da una molteplicità di fini e una elevata complessità (Filippucci e Marliani 1992b). L'indagine condotta attualmente dall'Istat è ripetuta e presenta numerose difficoltà che possono introdurre alti errori extracampionari: mancate collaborazioni di alcuni Comuni, carenza di professionalità nei rilevatori, scarsa supervisione della qualità del loro lavoro, eccetera (Filippucci e Marliani 1992a). Inoltre, i dati sul reddito delle famiglie spesso sono poco attendibili e accurati, scarsa attenzione è posta alla rilevazione del risparmio, anche se reddito e risparmio sono indispensabili per comprendere le scelte e il comportamento dei consumatori¹⁰. La revisione dell'indagine Istat non dovrebbe, dunque, trascurare la possibilità di misurare correttamente anche reddito e risparmio migliorando gli strumenti di rilevazione per ottenere dati validi, attendibili, e precisi (Filippucci e Marliani 1992a).

Un'indagine attuata con un campionamento diverso da quello continuo (rotato, panel, split panel) consente di utilizzare strategie di rilevazione più accurate sia per il reddito e il risparmio, sia per le spese di beni e servizi che non si verificano frequentemente nel tempo. Inoltre, l'indagine panel sui bilanci di famiglia faciliterebbe la raccolta dei dati nel tempo anche per le altre categorie di spesa (affitto, acqua, luce, gas, eccetera) e per le spese rare (mobili, televisori, elettrodomestici, auto, *roulottes*, eccetera). Tuttavia, bisogna anche ricordare che è necessario chiedere ai partecipanti la fatica di registrare le spese correnti per un periodo, ad esempio due settimane, ogni anno (Duncan, Juster, e Morgan 1984; Marbach 1988).

In base all'esame dei metodi di indagine, considerate l'ampiezza e la complessità del fenomeno consumo con le relative difficoltà di misura e

di modellazione (Sudman 1976; Sudman e Ferber 1979; De Stavola 1986; Blundell 1988; Deaton 1992; Filippucci e Marliani 1992b), possiamo qui concentrarci su alcune ipotesi, suscettibili di discussione e approfondimenti, rispetto al metodo di indagine da utilizzare nella rilevazione dei dati.

a) *Si adotta un'indagine rotata.*

In tal caso alcuni aspetti di interesse teorico, come la "verifica" della teoria del reddito permanente o l'esplorazione dei problemi connessi all'aggregazione (Deaton 1992), sono sacrificati perché i soggetti rimangono per poco tempo nel campione. Tuttavia, tale presenza, per quanto ridotta, permette: di migliorare la qualità dei dati raccolti; di controllare gli errori di memoria e telescopici in eventuali domande retrospettive; di attivare meccanismi per misurare reddito, risparmio, e loro impiego (investimento); di aumentare la raccolta delle spese rare e la precisione della loro misura; di mantenere la rappresentatività del campione.

b) *Si adotta un'indagine split panel.*

La soluzione è interessante e offre notevoli potenzialità perché, disponendo di una componente panel e una componente rotata o ripetuta, permette di affrontare quasi tutte le questioni che emergono nell'analisi di fenomeni che evolvono nel tempo.

Diversi problemi, comunque, permangono e alcuni di essi richiedono una soluzione in fase di progettazione. Per esempio, non bisogna trascurare le questioni derivanti dall'abbinamento delle unità che si potrebbe effettuare durante l'analisi (U.S. Bureau of the Census 1985; Deaton 1985; Barr e Turner 1990). Inoltre, molte indagini ripetute sul consumo delle famiglie non rilevano dati attendibili di reddito e risparmio; il consumo è spesso superiore al reddito; sicché in fase di abbinamento delle due indagini possono sorgere difficoltà per le differenze tra i dati riscontrati nelle due componenti. La componente continua (non panel), quindi, potrebbe essere attuata con un campione rotato perché consente una rilevazione più precisa di alcune grandezze. In alternativa, bisogna sforzarsi per rendere più accurata la rilevazione di reddito, consumo, e risparmio agendo su altri fattori come, per esempio, i rilevatori.

c) *Si adotta un'indagine panel.*

Il metodo dà una risposta adeguata alle esigenze, emerse nell'ultimo decennio, di ottenere dati che esprimano dinamiche individuali. Infatti, solo con i dati provenienti da un'indagine panel si possono introdurre nei modelli sul consumo i caratteri e le preferenze degli individui: sesso, età,

composizione familiare, titolo di studio, zona di residenza, occupazione, eccetera. Tali caratteri producono effetti che sono certamente correlati con il consumo e con il reddito (Deaton 1992).

Se l'indagine panel è condotta per un periodo sufficientemente lungo, aumentano le possibilità di verificare alcuni aspetti di politica economica (Blundell 1986, 1988), di esplorare gli effetti dell'aggregazione, di modellare la domanda di consumo di certi beni, e così via. La disponibilità di microdati, perciò, costituisce un importante fattore per l'approfondimento e la verifica della teoria del consumo (Deaton 1992; Filippucci e Marliani 1992b) che giustifica gli sforzi richiesti nella conduzione dell'indagine panel e l'opportunità di adottarla per rilevare il consumo delle famiglie.

Si è detto che la realizzazione di un'indagine panel necessita di una organizzazione complessa e efficiente. Per determinare la struttura che dovrebbe curarne lo svolgimento è necessario approntare un'indagine sperimentale e limitare il territorio di riferimento a tre o quattro regioni. L'indagine panel pilota, infatti, consente di: verificare la sua fattibilità; sviluppare competenze e esperienza; stabilire un inventario di informazioni statistiche necessarie; appurare se la struttura operativa di lavoro è efficace; controllare se le fasi della rilevazione sono state programmate adeguatamente, dal campionamento all'intervista, alla "convalidazione" delle misure, all'elaborazione dei dati, all'analisi. Tre occasioni o fasi potrebbero rappresentare il numero minimo per tentare di organizzare l'indagine panel nei dettagli e per costituire una base di esperienze sul campo inerenti alle strategie di rilevazione, alle difficoltà del questionario, al controllo della qualità e all'elaborazione statistica dei dati.

Nel caso dell'indagine sui consumi e sui bilanci delle famiglie emergono ulteriori complicazioni e difficoltà sia per il rispondente, sia per l'intervistatore (Filippucci e Marliani 1992b); è probabile, quindi, che da un'indagine pilota emerga la necessità di adottare il metodo dell'indagine ripetuta per la rilevazione delle spese correnti. Tra tutte le cause che potrebbero influenzare tale esito, si può citare la fatica e il notevole impegno necessari per compilare i moduli e registrare le spese personali e della famiglia; , er tanto, non si può chiedere alle famiglie di restare a lungo nel campione senza amplificare tutti gli svantaggi dell'indagine panel.

Un'ultima considerazione scaturisce dalla complessità e dalla molteplicità dei fini dell'indagine sui consumi. Infatti, tali aspetti favoriscono probabilmente l'uso di una combinazione di metodi di indagine perché, come

si è visto, ogni metodo ha potenzialità differenti, sicché bisogna definire che cosa si intende indagare e, quindi, quali siano i metodi più adeguati. Per esempio, le spese correnti si possono rilevare con le indagini ripetute o rotate; le spese sanitarie e/o le spese per i beni durevoli e semidurevoli si possono raccogliere aggiungendo domande retrospettive al questionario utilizzato per la rilevazione delle spese correnti e/o tramite un'indagine panel. Si ottiene, così, un'indagine *split panel* che attribuisce obiettivi di rilevazione diversi alle due componenti, quella ripetuta o rotata e quella panel, e che richiede la soluzione di alcuni problemi tra cui quelli inerenti al possibile abbinamento delle unità in fase di analisi.

3. Aspetti caratteristici di un piano di campionamento per la rilevazione dei consumi: l'indagine panel e rotata

La scelta di un'indagine panel per la rilevazione dei consumi e dei bilanci delle famiglie richiede ovviamente di definire la popolazione innanzitutto e quali unità statistiche rilevare, le famiglie o gli individui o le abitazioni; di stabilire con quale periodicità eseguire la rilevazione, biennale o annuale o semestrale; di fissare per quante fasi si ripete l'intervista alle unità del campione. Questi aspetti devono essere definiti in anticipo perché connessi agli obiettivi e all'organizzazione dell'indagine. Così, per esempio, se si vuole verificare come muta il comportamento nell'acquisto di beni semidurevoli e durevoli, il contatto con le famiglie deve avvenire una volta ogni periodo di tempo dato; ma per la sua scelta non esistono criteri univoci e dipende sia dai fattori che determinano il fenomeno oggetto di studio, sia dalla tecnica di misurazione: domande retrospettive o compilazione di moduli lasciati presso l'intervistato. Il contatto delle famiglie, probabilmente, deve essere abbastanza frequente e bisogna appurare il modo più efficace di registrare le spese al fine di ottimizzare i costi dell'intervista. Per la periodicità della rilevazione e per una breve panoramica di indagini svolte in quest'ambito, con i relativi tratti distintivi, si veda Marbach (1988).

Alcuni elementi essenziali su cui ci soffermeremo, che caratterizzano i piani di campionamento per la rilevazione dei consumi e che possono differire da paese a paese, sono: un periodo fisso di rilevazione delle spese che varia da una settimana a un mese e a volte anche per categorie di beni, la distinzione tra città grandi e piccole o una stratificazione territoriale (unità di

primo stadio), la periodicità della rilevazione, il metodo di indagine (panel e rotata).

3.1. Periodo di registrazione delle spese

La registrazione giornaliera delle spese correnti viene effettuata per un periodo di tempo fissato, la cui ampiezza non è determinata secondo regole precise; pertanto essa varia da paese a paese (10, 14, 30 giorni o 1 mese). Per la rilevazione dei consumi e dei bilanci di famiglia, l'EUROSTAT ha emanato delle direttive che fissano il periodo a 14 giorni, coincidenti con due settimane (Innocenzi 1992). Rispetto all'attuale disegno campionario, si ha così una riduzione del numero di periodi di registrazione all'anno: da 36 derivanti dalla decade a 26 perché tante sono le quattordicine in un anno. Se resta invariato lo schema di campionamento, ne consegue anche una riduzione del numero famiglie nel campione dell'ordine del 27-28%, cioè si passerebbe da circa 39000 a circa 25000 famiglie all'anno. Nel seguito si farà riferimento a un periodo di registrazione uguale a 14 giorni.

3.2. Stratificazione delle unità di primo stadio

In base alle esperienze accumulate dalle indagini svolte, anche in altri paesi, sembra opportuno creare raggruppamenti specifici per le unità di primo stadio concepibili come aree territoriali omogenee rispetto al consumo, anche se esistono diverse difficoltà di tipo pratico e amministrativo. Pertanto, è prioritario affrontare la ricerca di aggregazioni di aree geografiche che mostrano omogeneità nel comportamento dei consumatori al fine di costruire una base per il campionamento areale (Dippo e Jacobs 1983; U.S. Bureau of the Census 1985; Filippucci e Marliani 1992b). La stratificazione delle unità di primo stadio, verosimilmente, si dovrà basare, nei prossimi anni, ancora sulla distinzione tra Comuni autorappresentativi (AR) e non autorappresentativi (NAR), come nel piano di campionamento attuale. D'altra parte, tale schema viene adottato, in forme analoghe o appena variate, anche nelle indagini svolte in altri paesi. La stratificazione dei Comuni NAR in base all'altezza sul livello del mare e all'attività economica prevalente, invece, richiede ulteriori studi e riflessioni¹¹.

3.3. L'indagine panel limitata nel tempo

L'indagine panel, così com'è definita, non è sembra utilizzabile nell'indagine sui consumi perché la collaborazione risulta onerosa per gli intervistati e i limiti che la caratterizzano possono rendere inattendibili i dati raccolti. Infatti, in molti paesi il tasso di non risposta supera il 30% anche nelle indagini ripetute e risulta più alto dei tassi di non risposta osservati nelle altre indagini sulle famiglie (Innocenzi 1992). In Italia il tasso di non risposta è più basso, ma non esistono esperienze sul mantenimento dei soggetti nel campione per un lungo periodo o sull'indagine panel. Tuttavia, se si sceglie un periodo breve di osservazione dei soggetti inclusi nel campione, per esempio ogni unità viene mantenuta per due o tre anni di seguito, allora l'indagine panel, che si dirà "limitata nel tempo", diventa applicabile con minori difficoltà alla rilevazione dei consumi e dei bilanci delle famiglie. Il metodo rappresenta un'esperienza interessante, anche se venisse applicato solo nella fase sperimentale, necessaria per verificare sul campo le sue potenzialità. La limitazione nel tempo indica che lo stesso campione viene mantenuto per due o tre anni, ma quattro o cinque anni sarebbero più efficaci per la modellazione dei dati: una famiglia intervistata in un dato periodo, perciò, viene intervistata di nuovo nello stesso periodo dell'anno successivo, e anche di due anni dopo quando il periodo di permanenza è di tre anni. Supponendo che nell'anno t si sia estratto un campione, c_t , di famiglie e sia stato ripartito tra le 26 quattordicine, q , lo schema è descritto simbolicamente nel prospetto 1 dove $(n_{jkrqt})_{c_t}$ indica la j -esima famiglia

Prospetto 1 — Schema di un campionamento panel limitato nel tempo: ogni campione "vive" tre anni.

$(n_{jkrqt})_{c_t}$	$(n_{jkrq,t+1})_{c_t}$	$(n_{jkrq,t+2})_{c_t}$
$(n_{lkrq,t+3})_{c_{t+3}}$	$(n_{lkrq,t+4})_{c_{t+3}}$	$(n_{lkrq,t+5})_{c_{t+3}}$
...

da intervistare nel k -esimo Comune della provincia r -esima, nella q -esima quattordicina dell'anno t del campione estratto in quell'anno c_t ; la stessa famiglia viene intervistata anche nella corrispondente quattordicina dell'anno $t+1$ e dell'anno $t+2$ se il periodo di permanenza è di tre anni. Nell'anno $t+3$ si estrae un nuovo campione di famiglie, c_{t+3} , le quali vengono intervistate negli anni $t+3$, $t+4$, e $t+5$, come indicato nel prospetto 1.

Si noti che l'indice j dipende da q perché le famiglie sono diverse da una quattordicina all'altra. Infatti, si supponga di estrarre un campione di famiglie, c_t , al tempo t pari a 25740; a ogni quattordicina dell'anno se ne intervistano 990 delle 25740. Negli anni successivi si intervistano le stesse famiglie nello stesso ordine temporale con il quale sono state intervistate l'anno precedente: le 990 famiglie intervistate nella prima quattordicina di gennaio dell'anno t vengono intervistate sempre nella prima quattordicina dell'anno $t+1$ e dell'anno $t+2$, le 990 famiglie intervistate nella seconda quattordicina di gennaio dell'anno t vengono intervistate sempre nella seconda quattordicina dell'anno $t+1$ e dell'anno $t+2$, e così via.

Le percentuali elevate di mancate collaborazioni osservate già nelle indagini ripetute, l'impegno notevole richiesto ai partecipanti in termini di tempo e di attenzione, e gli svantaggi propri dell'indagine panel (auto-selezione iniziale, attrito, reattività alla reintervista, eccetera) inducono a ipotizzare di non mantenere una famiglia per più di tre fasi nel campione. Tuttavia, occorrono studi specifici per comprendere quali sono i limiti di tempo ottimali entro i quali impegnare una famiglia nella rilevazione senza comprometterne la collaborazione.

Le indagini panel limitate nel tempo (reintervista delle unità dopo uno e, eventualmente, due anni), devono prevedere, tuttavia, accorgimenti particolari in sede di predisposizione del questionario per rilevare nell'anno successivo, con domande retrospettive, alcuni eventi accaduti nel periodo compreso tra le due interviste e ciò può favorire gli errori dovuti alle imprecisioni nei ricordi degli intervistati (Sudman e Brandburn 1973; Moss e Goldstein 1979; Cannell, Miller, e Oksenberg 1981; Mathiowetz e Duncan 1984).

Si potrebbe ipotizzare un ulteriore contatto a metà dell'anno per rilevare solo alcune spese particolari e rinforzare il rapporto dell'intervistatore con le famiglie, ma si può presumere che il costo organizzativo in questo caso diventi elevato e che richieda molta efficienza e capacità di gestione degli elenchi delle famiglie selezionate e delle diverse schede rilevate. I costi organizzativi aumentano, in particolare, perché il numero dei contatti con le famiglie, solo per la raccolta dati, raddoppiano e passano da 26 a 52 all'anno. Infatti, a regime, in ogni quattordicina q dell'anno t si devono intervistare le famiglie nel campione per la quattordicina corrente più le famiglie intervistate la tredicesima quattordicina antecedente. Lo schema è formalizzato nel prospetto 2, dove $(n_{lkr,(q+13)\bmod 26,t+[(q+13)/26]})$ indica le

Prospetto 2 — Schema di un campionamento panel limitato a tre anni con intervista breve delle unità dopo tredici quattordicine.

$(n_{jkrq,t})_{c_t}$	+	$(n_{lkr,(q+13)\bmod 26,t-1+\lfloor(q+13)/26\rfloor})_{c_{t+3(-1+\lfloor(q+13)/26\rfloor)}}$
$(n_{jkrq,t+1})_{c_t}$	+	$(n_{lkr,(q+13)\bmod 26,t+\lfloor(q+13)/26\rfloor})_{c_t}$
$(n_{jkrq,t+2})_{c_t}$	+	$(n_{lkr,(q+13)\bmod 26,t+1+\lfloor(q+13)/26\rfloor})_{c_t}$
$(n_{pkrq,t+3})_{c_{t+3}}$	+	$(n_{jkr,(q+13)\bmod 26,t-1+\lfloor(q+13)/26\rfloor})_{c_{t+3+3(-1+\lfloor(q+13)/26\rfloor)}}$
...	+	...

famiglie intervistate 13 quattordicine precedenti (circa sei mesi). Infatti, $(q+13)\bmod 26$ esprime il modulo o resto e $\lfloor(q+13)/26\rfloor$ il quoziente della divisione tra $(q+13)$ e 26. Così, per esempio, nella prima quattordicina di marzo, q circa uguale a 5, dell'anno 1992 si intervistano le 990 famiglie del campione corrente e si "rivisitano" le 990 famiglie intervistate nella tredicesima quattordicina antecedente che corrisponde circa alla prima quattordicina del mese di settembre ($q = 5 + 13 = 18$), dell'anno $1991 = 1992 - 1 + \lfloor(5+13)/26\rfloor$; nella seconda quattordicina di luglio ($q=15$) del 1992 si intervistano sempre le 990 famiglie del campione corrente e si "rivisitano" le 990 famiglie intervistate nella seconda quattordicina di gennaio ($q = 15 - 13 = 2$) dell'anno $1992 = 1992 - 1 + \lfloor(15+13)/26\rfloor$. L'indice $c_{t+3+3(-1+\lfloor(q+13)/26\rfloor)}$ specifica a quale campione panel appartiene la famiglia da intervistare. Quando si rinnova il panel nell'anno $t+3$, per esempio, nella prima quattordicina di gennaio si intervistano le 990 famiglie estratte per la prima volta e si "rivisitano" le 990 famiglie intervistate circa nella prima quattordicina di luglio appartenenti ancora al campione panel estratto al tempo t (tre anni prima); infatti, $t+3+3(-1+\lfloor(1+13)/26\rfloor) = t$. Il campione quattordicinale di 990 famiglie è tratto dall'esempio precedente. Aumenta, quindi, il lavoro degli intervistatori.

Nel prospetto 3 è riportato uno schema completo dall'inizio della costituzione del campione al suo rinnovamento in cui le 990 famiglie interessate in ogni quattordicina sono indicate con le lettere dell'alfabeto. A lettere e a simboli differenti (maiuscoli, minuscoli) corrispondono campioni differenti.

Prospetto 3 — Schema di un campione panel limitato a tre anni con "intervista breve" delle unità dopo 13 quattordicine.

Quatt.	t	$t+1$	$t+2$	$t+3$	$t+4$...
1	A	A N	A N	a N	a n	...
2	B	B O	B O	b O	b o	...
3	C	C P	C P	c P	c p	...
4	D	D Q	D Q	d Q	d q	...
5	E	E R	E R	e R	e r	...
6	F	F S	F S	f S	f s	...
7	G	G T	G T	g T	g t	...
8	H	H U	H U	h U	h u	...
9	I	I V	I V	i V	i v	...
10	J	J W	J W	j W	j w	...
11	K	K X	K X	k X	k x	...
12	L	L Y	L Y	l Y	l y	...
13	M	M Z	M Z	m Z	m z	...
14	N A	N A	N A	n a	n a	...
15	O B	O B	O B	o b	o b	...
16	P C	P C	P C	p c	p c	...
17	Q D	Q D	Q D	q d	q d	...
18	R E	R E	R E	r e	r e	...
19	S F	S F	S F	s f	s f	...
20	T G	T G	T G	t g	t g	...
21	U H	U H	U H	u h	u h	...
22	V I	V I	V I	v i	v i	...
23	W J	W J	W J	w j	w j	...
24	X K	X K	X K	x k	x k	...
25	Y L	Y L	Y L	y l	y l	...
26	Z M	Z M	Z M	z m	z m	...

3.4. L'indagine rotata

L'indagine rotata sembra apparentemente più complicata della precedente, sebbene, si possa semplificare qualora si esegua la rotazione sull'anno e la quota di permanenza sia abbastanza elevata, per esempio del 50%. In questo caso, a ogni periodo di registrazione della spesa, q , la metà delle famiglie nel campione è quella dell'anno precedente e l'altra metà è costituita da famiglie di nuova estrazione. Come è noto (Cochran, 1977; Kish, 1989), la quota di rotazione desiderabile è $P=1/3$; ma per semplificare, qualora in fase sperimentale risultasse che non è conveniente intervistare la stessa

famiglia più di due volte, si può fissare $P = 1/2$, sicché una famiglia resta per due fasi nel campione. Lo schema che segue, prospetto 4, "formalizza" quanto sopra è descritto.

Prospetto 4 — Schema di campionamento rotato su un periodo annuale con $P = 1/2$.

$(n_{jkrq,t-1}$	+	n_{lkrqt}
$(n_{lkrqt}$	+	$n_{mkrq,t+1}$
$(n_{mkrq,t+1}$	+	$n_{pkrq,t+2}$
...	+	...

In ogni quattordicina, il campione è composto da una quota di famiglie intervistate nell'anno precedente e una quota identica ($P = 50\%$) di famiglie estratte dalla popolazione nell'anno corrente. Nella corrispondente quattordicina dell'anno successivo, quest'ultima quota di famiglie viene ancora intervistata e l'altra quota viene sostituita con un campione estratto nuovamente dalla popolazione. Questo schema consente di migliorare la qualità dei dati e di analizzare i cambiamenti individuali da un anno all'altro sulla base di un campione "dimezzato", ma comporta un aumento dei costi e un periodo breve di permanenza delle famiglie nel campione; pertanto, bisogna valutare sul campo o in fase sperimentale se è conveniente utilizzarlo nella rilevazione dei consumi.

Nel caso di una rotazione semestrale, risulta conveniente e pratico anche $P = 2/3$ perché così una famiglia viene intervistata anche a distanza di un anno. Infatti, dopo sei mesi una quota pari a due terzi delle famiglie del campione viene intervistata di nuovo, dopo dodici mesi, infine, solo un terzo delle famiglie resta ancora nel campione e viene perciò intervistato. A regime, quindi, ogni unità statistica viene intervistata 3 volte, a 6 mesi di distanza ciascuna, dando origine allo schema riportato nel prospetto 5.

Nel k -esimo Comune della r -esima regione il campione della q -esima quattordicina al tempo t è composto da tre quote uguali di famiglie. Una di esse è la nuova quota estratta dalla popolazione, indicata con n_{lkrqt} . L'altra si riferisce alla quota intervistata 13 quattordicine precedenti (circa sei mesi prima) e indicata con $n_{jkr,(q+13)\bmod 26,t-1+[(q+13)/26]}$. Infine, il rimanente un terzo, indicato con $n_{hkrq,t-1}$, è la quota di famiglie intervistate 26 quattordicine precedenti, ovvero l'anno prima circa (c'è lo sfasamento di

un giorno tra le quattordicine e l'anno solare).

Prospetto 5 — Schema di campionamento rotato ogni semestre con $P = 2/3$.

$n_{hkrq,t-1}$	+	$n_{jkr,(q+13)\bmod 26,t-1+[(q+13)/26]}$	+	n_{lkrqt}
$n_{jkr,(q+13)\bmod 26,t-1+[(q+13)/26]}$	+	n_{lkrqt}	+	$n_{pkr,(q+13)\bmod 26,t+[(q+13)/26]}$
n_{lkrqt}	+	$n_{pkr,(q+13)\bmod 26,t+[(q+13)/26]}$	+	$n_{skr,q,t+1}$
$n_{pkr,(q+13)\bmod 26,t+[(q+13)/26]}$	+	$n_{skr,q,t+1}$	+	$n_{ukr,(q+13)\bmod 26,t+1+[(q+13)/26]}$
$n_{skr,q,t+1}$	+	$n_{ukr,(q+13)\bmod 26,t+1+[(q+13)/26]}$	+	$n_{vkr,q,t+2}$
$n_{ukr,(q+13)\bmod 26,t+1+[(q+13)/26]}$	+	$n_{vkr,q,t+2}$	+	$n_{zkr,(q+13)\bmod 26,t+2+[(q+13)/26]}$
...	+	...	+	...

Un esempio chiarisce meglio il significato della simbologia. Il campione della 15^a quattordicina del 1992, corrispondente circa alla seconda quattordicina del mese di luglio, è composto da 990 famiglie (tanto per fissare un numero che è anche multiplo di 3). Un numero di 330 famiglie sono state estratte 26 quattordicine precedenti, cioè la seconda quattordicina del mese di luglio dell'anno precedente, $n_{...15,1991}$; esse sono state intervistate allora, 13 quattordicine dopo, e vengono intervistate anche in quella corrente. Un numero di 330 famiglie sono state estratte e intervistate 13 quattordicine precedenti, cioè $(15 + 13) \bmod 26$ che è uguale a 2 e corrisponde alla seconda quattordicina dell'anno 1992 = $(1992 - 1 + [(15 + 13)/26])$; esse vengono intervistate nella quattordicina corrente e anche dopo 13 quattordicine. Infine, 330 famiglie vengono estratte e intervistate per la prima volta. Il campione della 16^a quattordicina del 1992 è composto da 990 famiglie, tutte diverse da quelle precedenti, ma con una composizione analoga; e così via. Pertanto, nell'intervallo di un anno si trattano 26 campioni differenti di 990 elementi; ogni campione è composto da tre parti secondo lo schema indicato, che si può semplificare con la simbologia riportata nel prospetto 6.

La rotazione con una quota $P = 2/3$ può essere utilizzata anche quando il periodo di rotazione è annuale; così, a regime, ogni unità viene intervistata

tre anni di seguito. Nel caso in cui il periodo è semestrale non c'è coincidenza con gli anni solari e ciò complica gli schemi di stratificazione delle unità di primo stadio.

Prospetto 6 — Schema semplificato del prospetto 4.

$(a, b, c)_{qt}$	\rightarrow	$(b, c, d)_{(q+13)\bmod 26, t + \lfloor (q+13)/26 \rfloor}$	\rightarrow
$(c, d, e)_{q, t+1}$	\rightarrow	$(d, e, f)_{(q+13)\bmod 26, t+1 + \lfloor (q+13)/26 \rfloor}$	\rightarrow
...	\rightarrow	...	\rightarrow

Nell'indagine rotata, come si è detto, la permanenza di famiglie nel campione, è limitata; si perdono i vantaggi del panel, mentre si riduce la variabilità annuale del campione totale perché solo un mezzo (un terzo) delle famiglie intervistate in un anno viene rinnovato. Gli svantaggi potrebbero essere compensati da una rilevazione più accurata, ma rimane faticoso per l'intervistato sicché la percentuale delle non risposte può diventare alta.

3.5 La dinamica della popolazione obiettivo

L'utilizzo di un'indagine rotata o panel, anche limitata nel tempo, comporta l'esame delle modifiche che la popolazione subisce nel tempo: nuove unità si formano, altre scompaiono o si trasformano.

Le unità *entrano* nella popolazione quando nascono, immigrano, o raggiungono lo stato utilizzato per definirla; le unità *escono* dalla popolazione quando muoiono, emigrano, o perdono lo stato che li fa appartenere a essa. L'unità *i* sarà *esistente* al tempo t , N_{it} , se si è costituita in un periodo precedente a quello in corso.

L'unità *i* nasce al tempo t , B_{it} , quando, per esempio, i figli lasciano i genitori e costituiscono un nuovo nucleo familiare, quando due coniugi si separano o divorziano costituendo due nuclei familiari diversi da quello preesistente, quando due famiglie differenti (*single*, vedovi, divorziati) si uniscono formando un'unica famiglia e dissolvono le due precedenti, quando due soggetti costituiscono una coppia "di fatto" che però sfugge spesso alle liste anagrafiche.

L'unità *i* muore al tempo t , D_{it} , quando, per esempio, tutti i membri della famiglia originaria muoiono, quando due famiglie si uniscono in una

nuova, quando la famiglia si dissolve a causa di una separazione.

Nell'indagine panel, anche limitata nel tempo, si può supporre che le uscite e le trasformazioni osservate nel campione *riflettano* quanto si verifica nella popolazione (Hanefeld 1984; Klevmarken 1984).

Per rappresentare l'evoluzione nel tempo, della popolazione delle famiglie, bisogna individuare un meccanismo di inclusione dei nuovi ingressi (costituzioni) e delle immigrazioni. L'argomento è complesso e le soluzioni adottate sono numerose; tra gli altri, si vedano Narain, Kathuria, e Srivastava, (1987), Holt e Skinner (1989), Schiopu-Kratina e Srinath (1991). Per esempio, si riporta la seguente proposta (Lalla 1992a). Siano N_t il numero di unità della popolazione, $N_{t-1,t}$ il numero di unità esistenti al tempo $t-1$ e al tempo t , B_t il numero di unità non esistenti al tempo $t-1$ e presenti al tempo t (entrate).

- 1°. Al tempo $t = 1$ si seleziona il campione iniziale, di ampiezza pari a n_1 unità, dalla popolazione di N_1 individui.
- 2°. Al tempo t generico, $t > 1$, si includono gli elementi campionati al tempo $t-1$ nel campione attuale: siano $n_{t-1,t}$ le unità sopravvissute tra $t-1$ e t . Si aggiunge un nuovo campione a quello delle incluse e/o rimaste. Note le popolazioni sopravvissute, $N_{t-1,t}$, e entrate, B_t , al tempo t , si estrae un campione, b_t , solo dalla popolazione entrata, B_t . Il campione totale sarà: $n_t = n_{t-1,t} + b_t$. L'entità di b_t si può determinare dal rapporto: $(b_t/B_t) = (n_1/N_1)$. Se accade che $(n_t/N_t) < (n_1/N_1)$, allora si può decidere di "ripopolare" il campione estraendo $[(n_1/N_1) \cdot N_t] - n_t$ elementi dalla popolazione $(N_t - n_t)$, in base a criteri che tengano conto delle classi in cui vi siano state più perdite rispetto alla popolazione di riferimento.

Il numero di unità esistenti al tempo $t-1$ e scomparsi al tempo t , D_t (uscite), non è stato preso esplicitamente in considerazione in quanto per le unità sopravvissute si ha: $N_{t-1,t} = N_{t-1} - D_t$. Le unità morte nel campione, appartenendo alle uscite, *riflettono* o rappresentano, come si è detto, le unità morte della popolazione.

Per costituire un'indagine panel bisogna conservare nel campione le unità presenti il più a lungo possibile; per mantenere la sua rappresentatività rispetto alla popolazione corrente bisogna includere la dinamica che si è verificata in essa. Lo schema precedente permette di conseguire in modo soddisfacente le due finalità.

4. Metodi di analisi dei dati raccolti con indagini panel

I procedimenti di stima dei parametri, nei modelli utilizzati per l'analisi dei dati, dipendono dagli assunti adottati e, quindi, dal tipo di approccio all'inferenza statistica. Senza entrare nei dettagli¹², si ricordano l'approccio classico (*random based*) e l'approccio predittivo (*model based*). Nel primo, le distribuzioni di probabilità discendono dal meccanismo casuale con cui si selezionano le unità della popolazione. Nel secondo, si assume che la quantità misurata di una data grandezza, per esempio il consumo totale di una famiglia, Y_i , sia la realizzazione di una variabile casuale associata all' i -esima unità della popolazione (Royall 1970; Royall e Cumberland 1981, 1985). Quale sia l'approccio più adeguato nel contesto empirico è ancora oggetto di discussione, ma va sottolineato che l'indagine sul consumo è complessa e prevede una pluralità di obiettivi conoscitivi sicché il campione è probabilistico; infatti, la modellazione della popolazione di riferimento risulta difficile per la molteplicità di variabili e richiede un intervento incisivo dell'operatore che contrasta con il ruolo di un ente ufficiale, come l'Istat (Cicchitelli e Herzel 1989). Inoltre, nell'inferenza da popolazioni finite la scelta del piano di campionamento e l'individuazione dei metodi di stima costituiscono sempre un caso a sé che dipende sia dagli obiettivi dell'indagine, sia dalla conoscenza della popolazione.

In questo paragrafo ci si limiterà, comunque, a illustrare solo alcuni aspetti dei procedimenti di stima di modelli che utilizzano dati raccolti con l'indagine panel o rotata. In particolare, si esamineranno le stime di un parametro della popolazione (media o totale) e dei parametri di un modello che richiede o l'analisi multivariata "tradizionale" o l'analisi delle durate.

I procedimenti utilizzati per stimare i livelli dei caratteri, considerando due fasi successive dell'indagine, si collocano nell'alveo dell'approccio classico e sono indicati nel seguito con l'espressione "metodo campionario".

L'analisi multivariata, invece, e l'analisi delle durate comprendono diverse classi di modelli che consentono di trattare con efficacia i dati longitudinali. Tuttavia, la struttura del piano di campionamento viene spesso trascurata nelle applicazioni: la stima dei parametri dei modelli è ottenuta trattando i dati come provenienti da un campione casuale semplice e assumendo che gli errori si distribuiscano normalmente con media nulla e varianza incognita. Presumibilmente ciò accade perché è problematico modellare l'effetto di un piano di campionamento complesso, com'è quello per

la rilevazione del consumo; non esistono ancora metodologie standard; e le procedure non sono integrate negli strumenti di elaborazione più diffusi (Hidiroglou, Fuller, e Hickman 1980; Schnell e altri 1988). Si deve sottolineare, comunque, che l'esclusione o l'inclusione della strategia di campionamento adottata può influenzare i risultati e le stime; perciò, l'argomento dovrebbe essere ulteriormente approfondito con specifico riferimento ai modelli sul consumo. Alcune difficoltà tipiche, emergenti nell'impiego di tecniche di analisi multivariata, sono state evidenziate da Bebbington e Smith (1977) e Smith (1984).

Le indagini ripetute nel tempo, inoltre, danno origine a serie storiche sicché un altro approccio si basa su questa metodologia per ottenere previsioni e studiare la ciclicità dei fenomeni (Scott, Smith, e Jones 1977; Poli 1988; Cocchi 1989; Rao, Srinath, e Quenneville 1989). In quest'ambito è, invece, trascurata l'applicazione delle serie storiche.

4.1. Metodo campionario: stime di livello

Le stime della spesa totale, della spesa media mensile pro capite e per famiglia, della quantità consumata media mensile ottenute mediante l'indagine sui consumi delle famiglie viene qui esaminata e discussa solo per l'indagine rotata, l'estensione all'indagine panel è immediata.

I procedimenti di stima dei parametri della popolazione, studiati da diversi autori (Patterson 1950; Eckler 1955; Gourney e Daly 1965; Sen 1973; Graham 1973; Cochran 1977; Bellhouse 1991), sono noti e si tratta, perciò, di applicarli ai casi specifici sopra illustrati. Si suppone che il campione dell'indagine sui consumi sia a due stadi e che le unità di primo stadio, Comuni o aree territoriali omogenee rispetto al consumo, siano stratificate e suddivise per regioni. Sia \bar{y}_{2trhk} la media della spesa totale nel campione di unità, n_{2trhk} , selezionate per la prima volta nel Comune k dello strato h della regione r al tempo t . Sia \bar{y}_{1trhk} , la media della spesa totale nel campione di unità, n_{1trhk} , presenti al tempo t e al tempo $t-1$ nel Comune k dello strato h della regione r . Sia $n_{trhk} = n_{1trhk} + n_{2trhk}$ l'ampiezza totale del campione. La stima della media della spesa totale al tempo t , \bar{y}_{trhk} si ottiene, di solito, pesando le due stime "indipendenti", \bar{y}_{1trhk} e \bar{y}_{2trhk} , rispettivamente con l'inverso delle loro varianze, W_{1trhk} e W_{2trhk} ,

riportate nella terza e quarta riga della tavola 1 senza gli indici rhk :

$$\bar{y}'_{trhk} = \phi_{trhk} \bar{y}'_{1trhk} + (1 - \phi_{trhk}) \bar{y}'_{2trhk} \quad (1)$$

dove $\phi_{trhk} = W_{1trhk} / (W_{1trhk} + W_{2trhk})$. La varianza di \bar{y}'_{trhk} risulta:

$$V[\bar{y}'_{trhk}] = \frac{1}{W_{1trhk} + W_{2trhk}}. \quad (2)$$

Per $t = 2$ si ottiene il risultato relativo al campionamento eseguito in due fasi (Cochran, 1977, p. 346); le stime della media e le relative varianze sono riportate nelle prime due righe della tavola 1 senza gli indici rhk .

Tavola 1 — Stimatori e relative varianze per la parte fissa e la parte rotata nel campionamento in più fasi successive, non corrette per popolazioni finite.

n	Stimatore	Varianza	Peso
Parte fissa, n_1 , e parte rotata, n_2 , nei tempi $t = 1$ e $t = 2$			
n_1	$\bar{y}'_{12} = \bar{y}_{12} + \beta(\bar{y}'_1 - \bar{y}_{11})$	$\frac{S_2^2(1-\rho^2)}{n_1} + \rho^2 \frac{S_1^2}{n}$	$\frac{1}{W_{12}}$
n_2	$\bar{y}'_{22} = \bar{y}_{22}$	$\frac{S_2^2}{n_2}$	$\frac{1}{W_{22}}$
Parte fissa, n_1 , e parte rotata, n_2 , nel tempo t generico			
n_{1t}	$\bar{y}'_{1t} = \bar{y}_{1t} + \beta_t(\bar{y}'_{t-1} - \bar{y}_{1,t-1})$	$\frac{S_t^2(1-\rho_t^2)}{n_{1t}} + \rho_t^2 V(\bar{y}'_{t-1})$	$\frac{1}{W_{1t}}$
n_{2t}	$\bar{y}'_{2t} = \bar{y}_{2t}$	$\frac{S_t^2}{n_{2t}}$	$\frac{1}{W_{2t}}$

Per incorporare nelle stime dei parametri relative al tempo t le stime ottenute nei periodi precedenti si utilizza un modello lineare (Gourieroux e Roy 1978; Cocchi 1989) e si risolve con il metodo dei minimi quadrati generalizzati:

$$\begin{pmatrix} \bar{y}_{1,t-1,rhk} \\ \bar{y}_{1,trhk} \\ \bar{y}_{2,t-1,rhk} \\ \bar{y}_{2,trhk} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{y}'_{t-1,rhk} \\ \bar{y}'_{trhk} \end{pmatrix} + \begin{pmatrix} e_{t-1,rhk} \\ e_{t,rhk} \end{pmatrix} \quad (3)$$

che può essere riscritto, con un ovvio significato dei simboli,

$$y = X\beta + e \quad (4)$$

nella quale si assume l'ipotesi $E[e] = 0$ con varianza

$$V[e] = V = \sigma^2 \begin{pmatrix} 1/n_{1trhk} & \rho_{trhk}/n_{1trhk} & 0 & 0 \\ \rho_{trhk}/n_{1trhk} & 1/n_{1trhk} & 0 & 0 \\ 0 & 0 & 1/n_{2trhk} & 0 \\ 0 & 0 & 0 & 1/n_{2trhk} \end{pmatrix}.$$

La soluzione del vettore dei parametri è $\beta = (X'V^{-1}X)^{-1}X'V^{-1}y$, cioè

$$\begin{aligned} \hat{\beta}_{t-1,rhk} &= \hat{Y}_{t-1,rhk} = \bar{y}_{2,t-1,rhk} + \alpha_{trhk}(\bar{y}_{1,t-1,rhk} - \bar{y}_{2,t-1,rhk}) \\ &\quad + \alpha_{trhk} \rho_{trhk} \frac{n_{2trhk}}{n_{trhk}} (\bar{y}_{2,trhk} - \bar{y}_{1,trhk}) \\ \hat{\beta}_{trhk} &= \hat{Y}_{trhk} = \bar{y}_{2,trhk} + \alpha_{trhk}(\bar{y}_{1,trhk} - \bar{y}_{2,trhk}) \\ &\quad + \alpha_{trhk} \rho_{trhk} \frac{n_{2trhk}}{n_{trhk}} (\bar{y}_{2,t-1,rhk} - \bar{y}_{1,t-1,rhk}). \end{aligned} \quad (5)$$

dove $\alpha_{trhk} = n_{trhk} n_{1trhk} / (n_{trhk}^2 - \rho_{trhk}^2 n_{2trhk}^2)$. La varianza degli stimatori o "soluzioni" è $V[\hat{\beta}] = (X'V^{-1}X)^{-1}$, cioè

$$V[\hat{\beta}] = \sigma^2 \begin{pmatrix} \alpha_{trhk} & \rho_{trhk} n_{1trhk} / n_{trhk} \\ \rho_{trhk} n_{1trhk} / n_{trhk} & \alpha_{trhk} \end{pmatrix}. \quad (6)$$

Si noti che le soluzioni ottenute sono combinazioni lineari degli stimatori elementari, derivanti sia dalla parte fissa e sia dalla parte rotata nelle due fasi del tempo; esse dipendono, poi, dalla struttura di correlazione esistente nelle grandezze oggetto di studio. Inoltre, le espressioni risultanti, non solo si basano su assunti semplificatori del fenomeno, ma non si applicano facilmente perché lo stimatore $\hat{\beta}$ deve essere ricalcolato per ogni fase in cui si raccolgano nuovi dati.

4.2. Metodi di analisi multivariata

I modelli multivariati lineari generalizzati, GLMM (*Generalised Linear Multivariate Model*), rappresentano uno strumento utile per l'analisi di dati

che si ottengono in esperimenti o indagini in cui T misure vengono effettuate sulle n unità o soggetti osservati. Tali misure possono essere caratteri distinti o la stessa caratteristica misurata a T periodi diversi. Sia \mathbf{Y} la matrice dei dati ($n \times T$) dove ogni elemento y_{jt} denota la t -esima misura effettuata sul j -esimo soggetto. Il riferimento di base è il solito modello di regressione lineare:

$$\begin{aligned} E[\mathbf{Y}] &= \mathbf{XB} \\ V[\mathbf{Y}_j] &= \Sigma \end{aligned} \quad (7)$$

dove \mathbf{Y}_j è la riga j -esima di \mathbf{Y} ; \mathbf{X} ($n \times q$) è la matrice del modello; \mathbf{B} ($q \times T$) e Σ ($T \times T$) sono le matrici dei parametri incogniti del modello. Si assume che le grandezze sono vettori di osservazioni complete per ogni persona e la stessa matrice del disegno per ciascuna delle T rilevazioni. Un vettore unico di parametri, \mathbf{b}_t , viene stimato a ogni periodo generico, t , di rilevazione: $t = 1, \dots, T$. L'espressione esplicita della (7) nel caso più semplice è:

$$y_{jt} = \beta_{1jt} + \sum_{l=2}^q \beta_{ljt} x_{ljt} + e_{jt} \quad j = 1, \dots, n \quad t = 1, \dots, T. \quad (8)$$

Assumendo lo stesso β per tutti i soggetti j , si ha per $j = 1, \dots, n$:

$$\mathbf{y}_j = \mathbf{X}_j \beta + \mathbf{e}_j \quad (9)$$

cioè

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{pmatrix} \beta + \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_n \end{pmatrix} \quad (10)$$

dove ogni sottomatrice o sottovettore ha T osservazioni e gli assunti sono quelli ordinari:

$$\begin{aligned} E[\mathbf{e}] &= 0 \\ V[\mathbf{e}] &= \begin{bmatrix} \sigma_{11}\Omega_{11} & \sigma_{12}\Omega_{12} & \dots & \sigma_{1n}\Omega_{1n} \\ \sigma_{21}\Omega_{21} & \sigma_{22}\Omega_{22} & \dots & \sigma_{2n}\Omega_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1}\Omega_{n1} & \sigma_{n2}\Omega_{n2} & \dots & \sigma_{nn}\Omega_{nn} \end{bmatrix} \end{aligned} \quad (11)$$

dove Ω_{jl} è una matrice $T \times T$ per $\forall j, l$. Diversi modelli sono ottenuti variando la struttura di \mathbf{V} . Gli assunti della regressione classica specificano

$E[e_{jt}] = 0$, $V[e_{jt}] = \sigma^2$, $Cov(e_{jt}, e_{ls}) = 0$ per $t \neq s$ o $j \neq l$. Se si assume, invece, che c'è eteroschedasticità fra i soggetti, si ha:

$$V[\mathbf{e}] = \begin{bmatrix} \sigma_1 \mathbf{I} & 0 & \dots & 0 \\ 0 & \sigma_2 \mathbf{I} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \mathbf{I} \end{bmatrix}$$

Il metodo dei minimi quadrati generalizzati dà come risultato:

$$\hat{\beta} = [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}] = \left[\sum_{j=1}^n \frac{1}{\sigma_j^2} \mathbf{X}'_j \mathbf{X}_j \right]^{-1} \left[\sum_{j=1}^n \frac{1}{\sigma_j^2} \mathbf{X}'_j \mathbf{y}_j \right]. \quad (12)$$

Il modello lineare (8) può avere, quindi, diverse specificazioni che si suddividono tra coefficienti costanti e coefficienti variabili. Tra i modelli con coefficienti costanti si distinguono i casi: $\beta_{1jt} = \beta_1$ costanti per $\forall l$ e gli errori e_{jt} autocorrelati, β_{1jt} variabili e β_{ljt} costanti per $l \geq 2$. In quest'ultimo caso si può assumere che $\beta_{1jt} = \beta_{1j}$ e vari solo tra gli individui oppure che $\beta_{1jt} = \bar{\beta}_1 + \mu_j + \lambda_t$ e vari tra gli individui e i periodi di osservazione. In ognuno di questi modelli, i coefficienti possono essere costanti (modello a variabili di comodo) o casuali (modello a componenti degli errori). Una suddivisione analoga si ha nel caso di modelli a coefficienti variabili.

La formulazione generale dei GLMM può essere estesa per risolvere problemi differenti; si ottengono, così, numerosi modelli. L'analisi dell'argomento esorbita dagli scopi del presente lavoro; per un approfondimento si rinvia a Judge e altri (1985), De Stavola (1986), Hsiao (1986), Trivellato (1990), e Greene (1991).

4.9. Metodi di analisi delle durate

Nell'ultimo decennio l'analisi dei dati di sopravvivenza o delle "durate" ha suscitato un notevole interesse tra gli studiosi di aree diverse e, in particolare, tra gli economisti. La letteratura è vasta, sicché tra i numerosi autori si citano solo Cox e Oakes (1984), Kiefer (1988), Lancaster (1990), Torelli (1990). Particolare rilievo ha il volume 11 degli *Annali di statistica*, curato da Trivellato (1991), sia per il piano di campionamento, sia per i

modelli di durata, sia per le altre tecniche di analisi che si possono utilizzare in un'indagine complessa. Le indagini panel, *split* panel, e retrospettive generano dati che si possono trattare anche con questa classe di modelli, i quali partono generalmente dalla definizione di funzione di rischio:

$$h(t) = P[T \geq t] = -\frac{S'(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \quad (13)$$

dove $S(t) = P[T \geq t] = 1 - F(t)$ è la funzione di sopravvivenza che indica la probabilità di rimanere in uno stato (per esempio, disoccupato) fino a quando non si verifica un evento definito al tempo t (per esempio, assunzione); T è la variabile casuale che denota il tempo di permanenza in uno stato. La funzione di rischio esprime la probabilità "istantanea" di verificarsi di un evento (per esempio: trovare un posto di lavoro). La stima della funzione di rischio richiede di seguire nel tempo la "storia" degli individui oggetto di studio, sicché il metodo risulta spesso limitato dal breve periodo di osservazione che si ha che nelle indagini rotate. Inoltre, i modelli che si possono costruire contengono parametri e la loro stima richiede più assunti sulla funzione di sopravvivenza dei metodi non parametrici (del tipo stima limite-prodotto di Kaplan-Meier) o semiparametrici. Tuttavia, solo i metodi parametrici consentono un trattamento agevole dei fattori che determinano il processo in esame. Sia X_j un vettore di costanti relative a ogni j -esimo individuo che denoti un insieme di caratteri esplicativi o covariate coinvolte nel processo di durata. Il modello di rischio proporzionale è tra i più usati per verificare il loro effetto sulla funzione di sopravvivenza:

$$h_j(t; X) = h_{0j}(t)\psi(X; \beta) \quad (14)$$

dove $h_{0j}(t)$ è la funzione di rischio (*latente*) per il j -esimo individuo che ha il vettore $X_j = 0$. La funzione $\psi(X; \beta)$ viene spesso assunta uguale a un esponenziale, $\exp(X\beta)$, perché gli effetti delle covariate sulla funzione di rischio sono semplici da interpretare e il troncamento delle durate e altri casi di interruzione sono facili da trattare.

La stima del vettore dei parametri del modello può essere ottenuta con il metodo della massima verosimiglianza. Il valore dei parametri che determinano il punto di massimo del logaritmo della verosimiglianza vengono assunti come stime e si ottengono risolvendo il sistema di equazioni alle

derivate prime di

$$\begin{aligned} \log L(\theta, \beta) &= \sum_{j=1}^n d_{ij} \log [h_j(t; X)] + \sum_{j=1}^n \log [S_j(t; X)] \\ &= \sum_{j=1}^n \left\{ d_{ij} \log [h_{0j}(t; \theta)] + d_{ij} X_j \beta \right\} + \sum_{j=1}^n \log [S_j(t; X)] \end{aligned} \quad (15)$$

dove d_{ij} è una variabile binaria: $d_{ij} = 1$ se il periodo di durata del j -esimo individuo è completo o non interrotto al tempo t_i , $d_{ij} = 0$ se è ancora in corso. I vettori θ e β rappresentano i parametri della funzione di rischio latente e delle covariate, rispettivamente.

Il modello proporzionale discreto si applica sia quando la variabile casuale T è discreta in senso stretto, sia quando il periodo di tempo viene suddiviso in $I + 1$ intervalli disgiunti: $T_i = [t_{i-1}, t_i)$ per $i = 1, \dots, I$ con $t_0 = 0$ e $t_{I+1} = \infty$. Il tempo associato al fenomeno misurato coincide con l'intervallo di registrazione. Nel caso di un'indagine panel i periodi corrispondono ai giorni o alle quattordicine o ai mesi. Un approccio comune consiste nella sostituzione della funzione di rischio con un modello logistico discreto:

$$\frac{h_j(t; X)}{1 - h_j(t; X)} = \frac{h_{0j}(t; \theta)}{1 - h_{0j}(t; \theta)} \psi(X; \beta) \quad (16)$$

dove $h_j(t; X) = P[T \leq t + 1 | T > t]$ per il j -esimo individuo con vettore di covariate X . Siano t_1, \dots, t_I i tempi osservati di permanenza in uno stato: $t_1 < \dots < t_I$; sia m_i la molteplicità del tempo t_i per $i = 1, \dots, I$. La probabilità condizionale che gli elementi j_1, \dots, j_m subiscano un cambiamento dall'insieme di rischio $\mathcal{R}(t_j)$, dati tutti gli elementi a rischio da t_j in poi, è data da:

$$\frac{\psi(j_1) \dots \psi(j_m)}{\sum_{k \in s(j; m)} \psi(k_1) \dots \psi(k_m)} \quad (17)$$

dove $s(j; m)$ indica l'insieme di tutte le selezioni di $m = m_j$ soggetti dall'insieme dei soggetti a rischio $\mathcal{R}(t_j)$. L'espressione si semplifica se si assume la relazione log-lineare: $\psi(X; \beta) = \exp(X\beta)$. Il logaritmo della funzione di verosimiglianza risulta, allora,

$$\log L(\beta) = \sum_{j=1}^I \left[s_j \beta - \log \left(\sum_{k \in s(j; m)} \exp(s_{jk} \beta) \right) \right] \quad (18)$$

dove $s_j = x_{j1} + \dots + x_{jm}$ è la somma dei vettori x_j degli individui j che cambiano stato al tempo t_j .

Un approccio simile (Folsom, La Vange, e Williams 1989) definisce, così la funzione di rischio:

$$h_i(t_i; X) = 1 - (1 - h_i)^{\exp(X\beta)} \quad i = 1, \dots, I+1 \quad (19)$$

dove h_i è la funzione di rischio "latente" per $X=0$ nell' i -esimo periodo di tempo, cioè per l'intervallo T_i ; X è il vettore delle covariate; β è il vettore dei coefficienti delle covariate. Dato un campione di ampiezza n , osservato negli $I-1$ periodi, la stima dei coefficienti β può ottenersi con il metodo della massima verosimiglianza pesando opportunamente le osservazioni con i pesi campionari, w_j , per il j -esimo individuo. La funzione di verosimiglianza è data da:

$$L(\tau, \beta) = \sum_{j=1}^n w_j \sum_{i=1}^{I-1} \left\{ d_{ij} \log \{ 1 - \exp[-\exp(\tau_i + x'_{ij}\beta)] \} - (1 - d_{ij}) \exp(\tau_i + x'_{ij}\beta) \right\} \quad (20)$$

dove $\tau_i = \log[-\log(1 - h_i)]$ e d_{ij} , come in precedenza, vale 1 se l'unità cambia stato nell'intervallo T_i e vale 0 altrimenti.

Gli aspetti analitici della ricerca delle soluzioni e i metodi di verifica dell'adattamento dei modelli descritti, come pure gli altri argomenti dell'analisi delle durate, si tralasciano; per un approfondimento si può cominciare dai riferimenti bibliografici sopra citati.

5. Conclusioni

La revisione dell'indagine sui consumi e sui bilanci delle famiglie deve interessare, come si è detto, tutte le fasi del processo di rilevazione; per un'esposizione dettagliata si rimanda a Filippucci e Marliani (1992a, b). In questo lavoro si è cercato di evidenziare l'importanza di altri metodi di indagine, panel o rotata, e si sono delineati alcuni schemi utili, tra i tanti, applicabili alla raccolta dei dati del consumo. Da quanto precede si può concludere che l'indagine ripetuta è la più conveniente per rilevare le spese correnti perché non pone particolari problemi organizzativi e non

affatica o scoraggia i rispondenti. Al fine di disporre di informazioni, inerenti alcune caratteristiche, più precise e continuative nel tempo per la stessa unità statistica, come il reddito o le spese per l'affitto, per le vacanze, per il riscaldamento, per il telefono, ..., pare auspicabile utilizzare una indagine panel, realizzando così di fatto un'indagine *split panel*, i cui obiettivi devono essere definiti. Inoltre, bisogna designare le variabili da rilevare e i criteri per abbinare statisticamente le unità al fine di non vanificare la possibilità di abbinamento tra le due componenti (U.S. Bureau of the Census 1985; Barr e Turner 1990).

Una possibile alternativa potrebbe essere data da un'indagine panel limitata nel tempo anche per la rilevazione delle spese correnti. La scelta tra l'indagine ripetuta e l'indagine panel limitata nel tempo si può effettuare solo con una sperimentazione, magari circoscritta a poche regioni. La sperimentazione permetterebbe anche di formare il gruppo che dovrà seguire successivamente la realizzazione dell'indagine e costituisce un elemento essenziale di giudizio e valutazione sulle possibilità offerte dalle indagini alternative in termini di efficacia nella raccolta e nel controllo della qualità dei dati. Indagini focalizzate su temi specifici delle varie fasi della rilevazione sono necessarie per appurare le difficoltà o per individuare le soluzioni migliori, perché solo intervenendo su tutte le fasi e gli ostacoli pertinenti si può ridurre l'errore extra-campionario e migliorare la precisione delle stime.

Gli schemi di indagine proposti sono strettamente connessi al piano di campionamento delle unità di primo stadio che deve essere ancora ridefinito (Lalla 1992b). Un'analisi più estesa e approfondita dei dati disponibili dell'indagine attuale è, comunque, utile per comprendere i limiti e verificare le alternative allo schema di campionamento; per accertare gli effetti sulle stime dell'ampiezza demografica del Comune, della stratificazione in base all'altimetria e all'attività economica prevalente. Tuttavia, per condurre l'elaborazione di questi dati occorrono forze numerose e risorse cospicue.

La discussione precedente concerne solo alcuni aspetti dei metodi di indagine, alternativi a quella ripetuta. Le ipotesi e le proposte esaminate costituiscono la base per un approfondimento ulteriore della revisione dell'indagine sui consumi e sui bilanci di famiglia, dal punto di vista sia progettuale, sia esecutivo. In particolare, rappresentano il primo stadio per ideare prima, e realizzare poi, un'indagine sperimentale, perché solo con una sperimentazione accurata è possibile individuare una soluzione sufficientemente robusta e verificata sul piano teorico e pratico.

Note

¹ Gli obiettivi (a) e (b) sono diversi tra loro. Il primo, (a), si riferisce alla stima di un parametro in un periodo di tempo nel quale i cambiamenti sono trascurabili; per esempio, il consumo medio pro-capite mensile. Il secondo, (b), si riferisce alla stima di un parametro, in un periodo di tempo, ottenuta dalla media delle stime relative a diversi periodi; per esempio, il consumo medio pro-capite mensile ottenuto dalla media dei consumi pro-capite mensili di un anno, il consumo medio pro-capite trimestrale ottenuto dalla media dei relativi consumi medi pro-capite mensili, eccetera (Duncan e Kalton 1987).

² L'obiettivo (g) concerne la rilevazione della frequenza, della ripetizione, e della durata di eventi che possono accadere nel tempo. Per esempio: la proporzione di famiglie (frequenza) che ha acquistato beni semidurevoli nell'anno passato; il numero di volte (ripetizione) che si verifica l'acquisto di uno stesso bene semidurevole in un biennio e la durata media dell'uso di tale bene.

³ Il piano di campionamento può predisporre che alcune unità: primarie, secondarie, eccetera, siano incluse anche più volte, nelle diverse fasi dell'indagine; tuttavia, non può prevedere che le unità individuali, oggetto di misurazione (per esempio, le persone o le famiglie), rimangano nel campione per più di una fase.

⁴ La popolazione equivalente è un insieme di unità individuato da caratteristiche date: area geografica, nazionalità, residenza, limiti di età, e così via. La sua composizione può variare nel tempo sia per le nascite e i decessi, sia per le emigrazioni e le immigrazioni (Duncan e Kalton 1987).

⁵ La stima della variazione netta (c) può essere ottenuta senza la scomposizione tra le variazioni avvenute negli individui e le variazioni nella composizione della popolazione, come gli ingressi per nascita o immigrazione e le uscite per decessi o emigrazione. Le osservazioni di eventi accaduti in un periodo fissato (g), antecedente alla data della rilevazione, possono essere affette da incontrollabili errori di memoria e telescopici.

⁶ Sinonimi di indagine panel sono indagine prospettica (*prospective survey*) e indagine multifase (*multi-round survey*). I termini studio o indagine longitudinale, utilizzati da Fienberg e Picard (1980) e da Fienberg e Tanur (1986), sono preferibili nel contesto di una indagine trasversale nella quale i dati sono acquisiti retrospettivamente (*retrospective survey*), come suggeriscono Duncan e Kalton (1987) e Holt e Skinner (1989). In epidemiologia viene frequentemente usato il

termine *follow-up*.

⁷ Le stime dei parametri della popolazione, riferite a tempi distinti (a) e a periodi che comprendono più fasi espresse tramite valori medi nel tempo (b), richiedono l'introduzione di un meccanismo che incorpori i cambiamenti della popolazione avvenuti a causa di ingressi e uscite. Pertanto, l'indagine panel, così com'è stata definita, non permette di conseguire pienamente gli obiettivi (a) e (b). La cumulazione dei campioni nel tempo (f) può attuarsi solo in alcuni casi; per esempio, si possono cumulare tutte le unità che hanno registrato un nuovo evento durante il periodo di osservazione, come i casi di divorzio o di occupazione.

⁸ L'indagine rotata si attua con un campione rotato (errato in italiano, anche se comune, ruotato). Il termine "rotato" è la traduzione del termine inglese *rotated* o *rotating*, proposto da Eckler (1955) per il campionamento in occasioni successive con sostituzione parziale delle unità già osservate. Le denominazioni variano tra gli autori: Patterson (1950) e Yeats (1949) lo chiamano proprio "campionamento in occasioni successive con sostituzione parziale delle unità", Hansen e altri (1955) "campionamento per una serie storica", Sukhatme e Sukhatme (1970) "campionamento consecutivo o successivo", Raj (1968) e Cochran (1977) "campionamento ripetuto o ripetitivo" (Fabbris, 1989). Duncan e Kalton (1987) usano il termine "indagine panel rotante". Si è preferito "indagine rotata" perché richiede un campione rotato e risulta una denominazione semplice e concisa.

⁹ La distorsione è incrementata anche dalle unità che entrano nella popolazione, appartengono a essa per un certo tempo, e escono prima della data dell'indagine.

¹⁰ Deaton (1992) sottolinea che anche l'indagine *Panel Study of Income Dynamics*, contenente dati sul reddito inerenti a un panel di famiglie dal 1967, non riporta informazioni sul risparmio e sul consumo totale.

¹¹ Sembra consigliabile mantenere la stratificazione altimetrica (montagna, collina, pianura) dei Comuni NAR in quanto si è notato in alcuni strati un certo guadagno in precisione nelle stime della spesa totale, della spesa media mensile (nell'anno) pro-capite e per famiglia, della quantità consumata media mensile (nell'anno) pro-capite (Russo, Falorsi, e Falorsi 1993). Il guadagno varia per tipo di bene e, sebbene rimanga prevalentemente intorno a cifre piccole 1-5%, in alcuni casi supera il 10%. Tuttavia, tale evidenza potrebbe dipendere ancora dall'ampiezza demografica del Comune perché la concentrazione dei Comuni piccoli varia secondo l'altitudine. Gli strati costruiti in base all'attività economica prevalente,

invece, risultano poco influenti sulle determinanti del consumo (Filippucci e Marliani 1992a). Per una discussione sugli schemi di stratificazione delle unità di primo stadio si veda Lalla (1992b).

¹² Un'esposizione esauriente degli aspetti essenziali e delle problematiche concernenti i diversi approcci all'inferenza statistica si può trovare, tra gli altri, in Cassel, Särndal, e Wretman (1977), Smith (184), Chaudhuri e Vos (1988), Cicchitelli e Herzel (1989), Cicchitelli, Herzel, e Montanari (1992).

Riferimenti bibliografici

- ALLISON P. (1981), *Discrete-time methods for the analysis of event histories*, in Leinhardt S. (ed.), "Sociological Methodology", pp. 61-98, San Francisco, Jossey-Bass.
- BARR S. R. e TURNER J. S. (1990), *Quality issues and evidence in statistical file merging*, in Liepins G. E. e Uppuluri V. R. R. (eds.), "Data Quality Control. Theory and Pragmatics", pp. 245-313, New York, Marcel Dekker.
- BEBBINGTON A. C. e SMITH T. M. F. (1977), *The effect of survey design on multivariate analysis*, in O'Muircheartaigh C. A. e Payne C. (eds.), "The analysis of survey data. Model fitting", Vol. 2, pp. 175-192, New York, John Wiley & Sons.
- BELLHOUSE D. R. (1991), *Marginal and approximate conditional likelihoods for sampling on successive occasions*, in "Survey Methodology", 17, 1, pp. 69-78.
- BLUNDELL R. (1986), *Econometric approaches to the specification of life-cycle labour supply and commodity demand behaviour*, in "Econometric Reviews", 5, pp. 89-170.
- BLUNDELL R. (1988), *Consumer behaviour: theory and empirical evidence — a survey*, in "Economic Journal", 98, N. 289, pp. 16-65.
- CANNELL C. F., MILLER P. V., e OKSENBERG L. (1981), *Research on interviewing techniques*, in Leinhardt S. (ed.), "Sociological Methodology", pp. 389-437, San Francisco, Jossey-Bass.
- CANTOR D. (1989), *Substantive implications of longitudinal design features: the National Crime Survey as a case study*, in Kasprzyk D., Duncan G. J., Kalton G., e Singh M. P. (eds.), "Panel Surveys", pp. 25-51, New York, John Wiley & Sons.
- CASSEL C. M., SÄRNDAL C. E., e WRETMAN J. H. (1977), *Foundations of Inference in Survey Sampling*, New York, Wiley & Sons.

- CHAUDHURI A. e VOS J. W. E. (1988), *Unified Theory and Strategies of Survey Sampling*, Amsterdam, North-Holland.
- CICCHITELLI G. e HERZEL A. (1989), *Campionamento da popolazioni finite: stato della ricerca e prospettive*, in "Atti del forum sul campionamento da popolazioni finite", pp. 7-51, Perugia, S.I.S. — Dipartimento di Scienze Statistiche.
- CICCHITELLI G., HERZEL A., e MONTANARI G. E. (1992), *Il campionamento statistico*, Bologna, il Mulino.
- COCCHI D. (1989), *Inferenza in campioni ruotati*, in "Atti del forum sul campionamento da popolazioni finite", pp. 53-68, Perugia, S.I.S. — Dipartimento di Scienze Statistiche.
- COCHRAN W. G. (1977), *Sampling Techniques*, third edition, New York, John Wiley & Sons.
- COX D. R. e D. OAKES (1984), *Analysis of Survival Data*, London, Chapman and Hall.
- DEATON A. S. (1985), *Panel data from time series of cross-section*, in "Journal of Econometrics", 30, pp. 109-126.
- DEATON A. S. (1992), *The microeconomics and macroeconomics of the permanent income hypothesis*, in "Temi di discussione", N. 166, Roma, Banca d'Italia.
- DE STAVOLA B. L. (1986), *Sampling designs for short panel data*, in "Econometrica", 54, 415-424.
- DIPPO C. S. e JACOBS C. A. (1983), *Area sample redesign for the consumer price index*, in "Proceedings of the Section on Survey Research Methods", American Statistical Association, pp. 118-123.
- DORMONT B. (1989), *Petite apologie des données de panel*, in "Economie et Prevision", 87, pp. 19-32.
- DUNCAN G. J. e HILL M. (1985), *Conceptions of longitudinal households: fertile or futile?*, in "Journal of Economic and Social Measurement", 12, pp. 361-375.
- DUNCAN G. J., JUSTER F. T., e MORGAN J. N. (1984), *The role of panel studies in a world of scarce resources*, in Sudman S. e Spaeth M. A. (eds.), "The Collection and Analysis of Economic and Consumer Behavior Data", pp. 301-328, Illinois, Champaign, Bureau of Economic and Business Research.
- DUNCAN G. J. e KALTON G. (1987), *Issue of design and analysis of surveys across time*, in "International Statistical Review", 55, pp. 97-117.

- ECKLER A. R. (1955), *Rotation sampling*, in "Annals of Mathematical Statistics", 26, pp. 664-685.
- FABBRIS L. (1989), *L'indagine campionaria. Metodi, disegni e tecniche di campionamento*, Roma, NIS: La Nuova Italia Scientifica.
- FIENBERG S. e PICARD R. R. (1985), *Designing surveys for measuring change in categorical data structures over time*, in Krishnaiah P. R. (ed.), "Multivariate analysis", pp. 60-93, Amsterdam, North-Holland.
- FIENBERG S. e TANUR J. (1986), *The design and analysis of longitudinal surveys: controversies and issues of cost and continuity*, in Boruch R. e Pearson R. (eds.), "Survey Research Designs: Towards a Better Understanding of Their Costs and Benefits", pp. 60-93, New York, Springer-Verlag.
- FILIPPUCCI C. (1992), *Sviluppi della teoria del consumo e problemi di misura*, in "Atti della XXXVI Riunione Scientifica della S.I.S.", vol. 2, pp. 11-18, Roma, Cisu.
- FILIPPUCCI C. e MARLIANI G. (1992a), *Rapporto finale progetto CON.PRI.*, nota interna 6, Roma, Istat, febbraio.
- FILIPPUCCI C. e MARLIANI G. (1992b), *La misura del consumo delle famiglie: una riflessione a partire dall'esperienza italiana*, CON.PRI. Rapporto di ricerca N. 6, Bologna, Dipartimento di Scienze Statistiche.
- FOLSOM R., LA VANGE L., WILLIAMS R. L. (1989), *A probability sampling perspective on panel data analysis*, in Kasprzyk D., Duncan G. J., Kalton G., e Singh M. P. (eds.), "Panel Surveys", pp. 108-138, New York, John Wiley & Sons.
- GOURIEROUX C. e ROY G. (1978), *Enquête en deux vagues, renouvellement de l'échantillon*, in "Annales de l'INSEE", 29, pp. 115-134.
- GRAHAM J. E. (1973), *Composite estimation in two cycle rotation sampling designs*, in "Communications in Statistics", 1, pp. 419-431.
- GREENE H. G. (1991), *Econometric Analysis*, New York, Macmillan.
- GURNEY M. e DALY J. F. (1965), *A multivariate approach to estimation in periodic sample survey*, in "Proceedings of the Social Statistics Section", American Statistical Association, pp. 242-257.
- HANSEN M. H., HURWITZ W. N., NISSELSOHN H., e STEINBERG J. (1955), *The redesign of the census current population survey*, in "Journal of the American Statistical Association", 50, pp. 701-719.
- HANEFELD U. (1984), *The German socio-economic panel*, in "Proceedings of the Social Statistics Section", American Statistical Association, pp. 117-124.

- HIDIROGLOU M. A., FULLER W. A., e HICKMAN R. D. (1980), *SUPER CARP*, Dept. of Statistics, Iowa, Ames, Iowa State University Press.
- HOLT D. e SKINNER C. (1989), *Components of change in repeated surveys*, in "International Statistical Review", 57, pp. 1-18.
- HSIAO C. (1986), *Analysis of Panel Data*, Cambridge, Cambridge University Press.
- INNOCENZI G. (1992), *Principali aspetti dell'indagine Istat sui consumi delle famiglie*, CON.PRI. Rapporto di ricerca N. 2, Bologna, Dipartimento di Scienze Statistiche.
- JESSEN R. J. (1942), *Statistical investigation of a sample survey for obtaining farm facts*, in "Iowa Agricultural Experimental Research Bulletin", 304, pp. 54-59.
- JUDGE G., HILL C., GRIFFITHS W., LEE T., e LUTKEPOHL H. (1985), *The Theory and Practice of Econometrics*, New York, John Wiley & Sons.
- KASPRZYK D., DUNCAN G. J., KALTON G., e SINGH M. P. (1989), *Panel Surveys*, New York, John Wiley & Sons.
- KIEFER N. M. (1988), *Economic duration data and hazard functions*, in "Journal of Economic Literature", XXVI (June), pp. 646-679.
- KISH L. (1983), *Data collection for details over space and time*, in Wright T. (ed.), "Statistical Methods and the Improvement of Data Quality", pp. 73-84, New York, Academic Press.
- KISH L. (1986), *Timing of surveys for public policy*, in "The Australian Journal of Statistics", 28, pp. 1-12.
- KISH L. (1989), *Sampling Methods for Agricultural Surveys*, Roma, FAO Statistical Development Series, N. 3.
- KLEVMARKEN N. A. (1984), *Household market and nonmarket activities. The first year of a Swedish panel study*, in "Proceedings of the Section on Social Statistics", American Statistical Association, pp. 125-131.
- LALLA M. (1992a), *L'indagine panel con rotazione su ingressi e uscite: alcune simulazioni sui dati INPS*, in "Atti della XXXVI Riunione Scientifica della S.I.S.", vol. 2, pp. 349-355, Roma, Cisu.
- LALLA M. (1992b), *Il piano di campionamento dell'indagine sui consumi: alcune proposte di variazione*, comunicazione al seminario: "Giornate di Statistica Economica 1992", Bressanone-Brixen, 16-18 settembre 1992.
- LANCASTER T. (1990), *The Econometric Analysis of Transition Data*, Cambridge, Cambridge University Press.

- LAZARSFELD P. F. e FISKE M. (1938), *The panel as a new tool for measuring opinion*, in "Public Opinion Quarterly", 2, pp. 596-612.
- LIPPI M. (1985), *Sulla dinamica delle relazioni tra variabili aggregate*, in "Politica Economica", I, N. 2, pp. 141-166.
- MANTEGAZZA S. e TASSINARI G. (1992), *La coerenza della valutazioni dei consumi privati nella contabilità nazionale e nell'indagine sui bilanci delle famiglie*, CON.PRI. Rapporto di ricerca N. 4, Bologna, Dipartimento di Scienze Statistiche.
- MARBACH G. (1988), *Le ricerche di mercato*, Torino, UTET.
- MATHIOWETZ N. e DUNCAN G. (1984), *Temporal patterns of response errors in retrospective reports of unemployment and occupation*, in "Proceedings of the Section on Survey Research Methods", American Statistical Association, pp. 652-657.
- MOSS L. e GOLDSTEIN H. (1979), *The Recall Method in Social Surveys*, London, University of London, Institute of Education.
- NARAIN P., KATHURIA O. P., e SRIVASTAVA A. K. (1987), *Measuring change in agricultural surveys*, in "Proceedings of the 46th Session of the ISI", IP-8.1, pp. 427-441.
- PATTERSON H. D. (1950), *Sampling on successive occasions with partial replacement of units*, in "Journal of the Royal Statistical Society", B, 12, pp. 241-255.
- POLI I. (1988), *A dynamic model for repeated sample surveys*, comunicazione a "International Dynamics Conference", California, S. Diego, 5-9 luglio 1988.
- RAJ D. (1968), *Sampling Theory*, London, McGraw-Hill.
- RAO J. N. K. e GRAHAM J. E. (1964), *Rotation designs for sampling on repeated occasions*, in "Journal of the American Statistical Association", 59, pp. 492-509.
- RAO J. N. K., SRINATH K. P., e QUENNEVILLE B. (1989), *Estimation of level and change using current preliminary data*, in Kasprzyk D., Duncan G. J., Kalton G., e Singh M. P. (eds.), "Panel Surveys", pp. 457-479, New York, John Wiley & Sons.
- ROYALL M. R. (1970), *On finite population sampling theory under certain linear regression models*, in "Biometrika", 57, pp. 377-387.
- ROYALL M. R. e CUMBERLAND W. G. (1981), *An empirical study of the ratio estimation and estimators of its variance*, in "Journal of the American Statistical Association", 76, pp. 67-77.
- ROYALL M. R. e CUMBERLAND W. G. (1985), *Conditional coverage properties of finite population confidence intervals*, in "Journal of the American Statistical Association", 80, pp. 355-359.
- RUSSO A., FALORSI P., e FALORSI S. (1993), *Indagine campionaria sui consumi delle famiglie. Strategie di campionamento e precisione delle stime*, CON.PRI. Rapporto di ricerca N. 3, Bologna, Dipartimento di Scienze Statistiche.
- SCHIOPU-KRATINA I. e SRINATH K. S. (1991), *Sample rotation and estimation in the survey of employment, payroll and hours*, in "Survey Methodology", 17, pp. 79-90.
- SCHNELL D., KENNEDY W. J., SULLIVAN G., PARK H. J., e FULLER W. A. (1988), *Personal computer variance software for complex survey*, in "Survey Methodology", 14, 1, 59-69.
- SCOTT A. J., SMITH T. F. M., e JONES R. G. (1977), *The application of time series methods to the analysis of repeated surveys*, in "International Statistical Review", 45, pp. 13-28.
- SEN A. R. (1973), *Theory and application of sampling on repeated occasions with several auxiliary variables*, in "Biometrics", 29, pp. 383-385.
- SILBERSTEIN A. R. (1973), *Recall effects in the U. S. consumer expenditure interview survey*, in "Journal of Official Statistics", 5, N. 2, pp. 125-142.
- SMITH T. M. F. (1984), *Sample surveys present position and potential developments: some personal views*, in "Journal of the Royal Statistical Society", A, 147, pp. 208-221.
- SOBOL M. G. (1959), *Panel mortality and panel bias*, in "Journal of the American Statistical Association", 54, pp. 52-68.
- SUDMAN S. (1976), *Applied Sampling*, New York, Academic Press.
- SUDMAN S. e BRANDBURN N. M. (1973), *Effects of time and memory factors on response in surveys*, in "Journal of the American Statistical Association", 68, pp. 805-815.
- SUDMAN S. e FERBER R. (1979), *Consumer Panels*, Chicago, American Marketing Association.
- SUKHATME P. V. e SUKHATME B. V. (1970), *Sampling Theory of Survey with Applications*, Second Edition, Iowa, Ames, Iowa State University Press.
- TORELLI N. (1990), *L'analisi dei dati di durata nelle scienze sociali: riflessioni metodologiche e una applicazione alla durata della disoccupazione*, in "Rivista di Statistica Applicata", 2, N. 1, pp. 7-26.

- TRIVELLATO U. (1990), *Modelli di comportamento e problemi di misura nelle scienze sociali: alcune riflessioni*, in "Atti della XXXV Riunione Scientifica della S.I.S.", vol. 1, pp. 11-34, Padova, Cedam.
- TRIVELLATO U. (a cura di) (1991), *Forze di lavoro: disegno dell'indagine e analisi strutturali*, "Annali di Statistica", anno 120, Serie IX, Vol. 11, Roma, ISTAT.
- TUMA N. e HANNAN M. (1984), *Social Dynamics: Models and Methods*, Florida, Orlando, Academic Press.
- U.S. BUREAU OF THE CENSUS e U.S. BUREAU OF THE LABOR STATISTICS (1985), *Proceedings of the Conference on Gross Flows in Labor Force Statistics*, Washington D.C.
- YATES F. (1949), *Sampling Methods for Census and Surveys*, London, Charles Griffin and Co. Ltd.